

MVP

Prediction of Appointment No Show

- **Baseline**

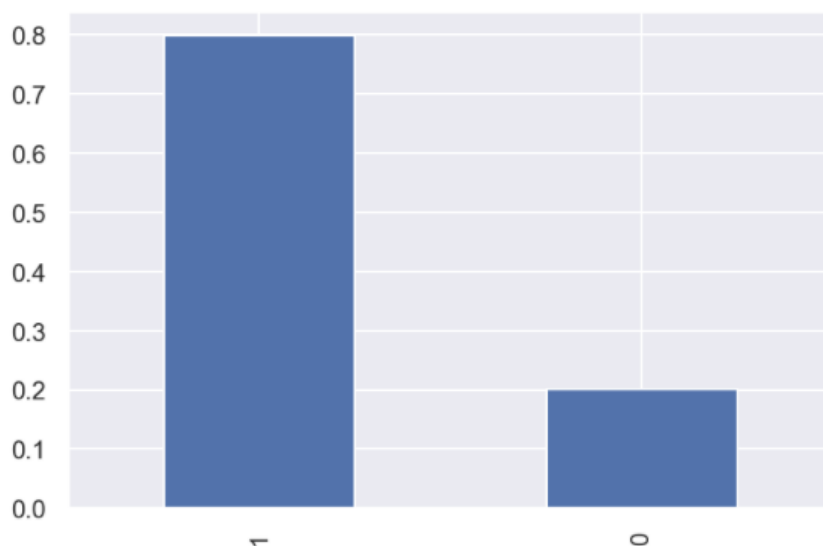
In this project there are 2 Logistic Regression BaseLine the first baseline made before oversampling and the other made after oversampling

- baseline before oversampling:
 - training score is 0.79828
 - test score is 0.79504
- baseline after oversampling:
 - training score is 0.5
 - test score is 0.796996

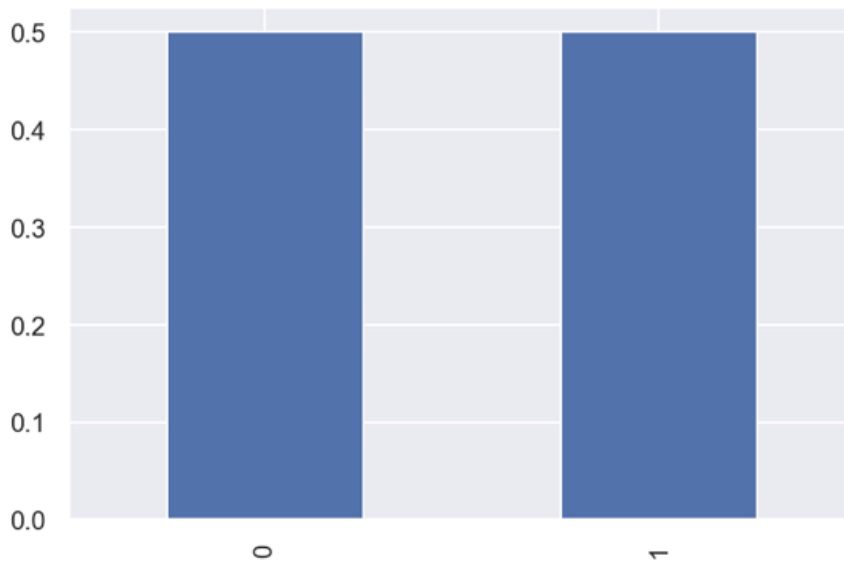
- **OverSampling**

in this section we used SMOT strategy to zoom the zero target which is 20% of the target while number of one's target equals 80% of whole the target

- **number of the target before OverSampling (80-20)**



- **number of the target after Over Sampling (50-50)**



- **Feature Correlation**

There is much correlation with the target (No-show), ScheduleMonth , AppointmentID have the strongest correlation with the target ,Also we have correlation with the age and while SMS-received has the weakest correlation with the target.

PatientId	1	0.004	0.0079	-0.0041	-0.0029	-0.0064	0.0016	0.011	-0.0079	-0.0097	0.0015	-0.0002	0.0016	0.0032	0.0024	0.0033	-0.0015
AppointmentID	0.004	1	0.018	-0.019	0.023	0.013	0.023	0.033	0.014	-0.26	0.16	0.029	0.49	0.019	0.82	-0.025	-0.77
Gender	0.0079	0.018	1	-0.11	-0.11	-0.056	-0.033	0.11	0.023	-0.046	0.0041	0.0034	-0.0061	0.00022	0.013	0.0028	-0.028
Age	-0.0041	-0.019	-0.11	1	-0.092	0.5	0.29	0.096	0.078	0.013	0.06	0.0035	0.015	-0.0071	-0.003	-0.0089	0.035
Scholarship	-0.0029	0.023	-0.11	-0.092	1	-0.02	-0.025	0.035	-0.0086	0.0012	-0.029	0.013	-0.0026	0.0026	0.017	0.0075	-0.03
Hipertension	-0.0064	0.013	-0.056	0.5	-0.02	1	0.43	0.088	0.08	-0.0063	0.036	0.033	0.0038	-0.0078	0.025	0.0011	-0.017
Diabetes	0.0016	0.023	-0.033	0.29	-0.025	0.43	1	0.018	0.058	-0.015	0.015	0.016	0.0037	-0.0042	0.026	-0.0071	-0.027
Alcoholism	0.011	0.033	0.11	0.096	0.035	0.088	0.018	1	0.0046	-0.026	0.0002	0.011	0.0039	-0.0025	0.029	-0.0043	-0.039
Handcap	-0.0079	0.014	0.023	0.078	-0.0086	0.08	0.058	0.0046	1	-0.024	0.0061	0.00033	0.0015	-0.0023	0.015	-0.0012	-0.02
SMS_received	-0.0097	-0.26	-0.046	0.013	0.0012	-0.0063	-0.015	-0.026	-0.024	1	-0.13	-0.05	0.11	-0.049	-0.29	0.17	0.4
No-show	0.0015	0.16	0.0041	0.06	-0.029	0.036	0.015	0.0002	0.0061	-0.13	1	0.043	0.021	-0.0038	0.16	-0.047	-0.19
Moring	-0.0002	0.029	0.0034	0.0035	0.013	0.033	0.016	0.011	0.00033	-0.05	0.043	1	0.015	-0.019	0.042	-0.012	-0.057
AppointmentMonth	0.0016	0.49	-0.0061	0.015	-0.0026	0.0038	0.0037	0.0039	-0.0015	0.11	0.021	0.015	1	-0.56	0.53	-0.21	0.0039
Appointment_day	0.0032	0.019	0.00022	0.0071	0.0026	-0.0078	-0.0042	-0.0025	-0.0023	-0.049	-0.0038	-0.019	-0.56	1	-0.15	0.37	0.048
ScheduledMonth	0.0024	0.82	0.013	-0.003	0.017	0.025	0.026	0.029	0.015	-0.29	0.16	0.042	0.53	-0.15	1	-0.45	-0.6
Scheduled_day	0.0033	-0.025	0.0028	-0.0089	0.0075	0.0011	-0.0071	-0.0043	-0.0012	0.17	-0.047	-0.012	-0.21	0.37	-0.45	1	0.042
days	-0.0015	-0.77	-0.028	0.035	-0.03	-0.017	-0.027	-0.039	-0.02	0.4	-0.19	-0.057	0.0039	0.048	-0.6	0.042	1
	PatientId	AppointmentID	Gender	Age	Scholarship	Hipertension	Diabetes	Alcoholism	Handcap	SMS_received	No-show	Moring	AppointmentMonth	Appointment_day	ScheduledMonth	Scheduled_day	days

- **Feature Engineering**

- Adding features
 - **Days:** the days between ScheduledDay and AppointmentDay the values of this feature are integer
 - **Morning:** the ScheduledDay is in the morning or not, the values of this feature are 0,1
 - **Appointment_Day**
 - **AppointmentMonth**
 - **Scheduled_Day**
 - **ScheduledMonth**
- Deleting useful features
 - **Neighborhood:** name of Neighborhood that has string values.
 - **ScheduledDay:** this feature has been converted to ScheduleDay and ScheduleMonth
 - **AppointmentDay:** this feature has been converted to Appointment_Day and AppointmentMonth

- **Models:**

- **Polynomial(degree = 2):** score = 0.2018
- **Polynomial(degree = 5):** score = 0.2030
- **Logistic:** test score = 0.7969
- **knn:** test score = 0.768
- **Random Forest:** test score = 0.7924
- **Extra Tree:** test score = 0.7836
- **Max voting:** test score = 0.7942
- **Average voting:** test score = 0.8015 "The best"
- **width voting:** test score = 0.8004
- **Scaler = 0.79441**
- **threshold:= 0.52**
- **Precision: 0.7970, Recall: 1.0000**
- **Threshold of 0.52:**
- **Precision: 0.7957, Recall: 0.5003**